Dictionaries/Lexicons Review

A lexicon is a list of words where their characteristics are annotated. Such characteristic may be Part of Speech, qualitative Sentiment Polarity, quantitative Sentiment Rating and many others.
Lexicons/dictionaries are used in Sentiment Analysis in order to give to each sentence its polarity from its very source: words and their sentence interaction. Obviously, since a sentence sentiment is not just the sum of each word own sentiment, a model that takes into account words interactions has to be built.
The main issue is that a lexicon is built for a certain context, therefore dictionaries are *context-oriented*. Since we are facing newspapers articles that deal with agricoltural commodities and previous sentiment research focused on products reviews, it appears clear that the most known lexicons are not oriented towards a newspapers articles context.
The aim of this research is to underline why a *review-oriented* dictionary will not produce accurate sentiment scores, then suggestions on how to build a new lexicon will be given and a list of other available lexicons will be provided.

Methodology:

The methodology will be a trivial one: a few phrases have been chosen from actual wheat-related articles, then *VADER Analysis Tool*, *Google Natural Language* sentiment analysis tool and *Linguistic Inquiry and Word Conts* results will be analyzed.
Analyzed sentences are listed below:

1. *July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006.*

2. *July wheat is expected to fall by nearly 14 percent to $4.00 per bushel, the lowest since September 2006.*

3. *July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006, while soybean November is seen tumbling 17 percent to $7.60 per bushel, the lowest since May 2007.*

4. *`Bladerunner` is often touted as one of the best science fiction films ever made. Indeed, it satisfies many of the requisites for good sci-fi: a future world with flying cars and humanoid robots attempting to rebel against their creators. But more than anything, `Bladerunner` is a fantastic exploration of the nature of what it means to be human. If we create robots which can think, will they become human? And if they do, what makes us unique? Indeed, how can we be sure we're not human in any case? `Bladerunner` explored these issues before such movies as `The Matrix,' and did so intelligently. The visual effects and score by Vangelis set the mood. See this movie in a dark theatre to appreciate it fully. Highly recommended!*

5. *What was Hollywood thinking with this movie! I hated, hated, hated it. BORING! I went afterwards and demanded my money back. They refused.*

Phrases 1, 2, 3 refer to price movements. In particular the pair 1, 2 are in economic terms at the opposites, while in phrase 3 wheat-related sentiment is polluted by soybean-related sentiment. Moreover in phrase 3 wheat and soybean sentiment *differ*, therefore there is an opposite sign pollution in wheat-related sentiment made by soybean-related sentiment.
These economic statements are characterised by a neutral lexicon with few adjectives used.
Phrases 4, 5 are Blade Runner movie reviews from the data given by Google NLP. They are opposite sign reviews and they will be used for testing purposes. It has to be noted that they strongly differ in the lexicon used when compared with the previous economic statements: the lexicon is colloquial, strongly subjective and many adjectives are used.

Assumed dictionaries behaviour is to evaluate differently all the economic phrases, namely 1, 2 and 3.

Google NLP and VADER results

Google NLP:

Phrase	***Result***
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006.	*polarity of -1 with magnitude of 0.3*
July wheat is expected to fall by nearly 14 percent to $4.00 per bushel, the lowest since September 2006.	*polarity of -1 with magnitude of 0.3*
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006, while soybean November is seen tumbling 17 percent to $7.60 per bushel, the lowest since May 2007.	*polarity of -1 with magnitude of 0.3*
`Bladerunner` is often touted as one of the best science fiction films ever made. Indeed, it satisfies many of the requisites for good sci-fi: a future world with flying cars and humanoid robots attempting to rebel against their creators. But more than anything, `Bladerunner` is a fantastic exploration of the nature of what it means to be human. If we create robots which can think, will they become human? And if they do, what makes us unique? Indeed, how can we be sure we're not human in any case? `Bladerunner` explored these issues before such movies as `The Matrix,' and did so intelligently. The visual effects and score by Vangelis set the mood. See this movie in a dark theatre to appreciate it fully. Highly recommended!	*polarity of 0.9 with magnitude of 5.3*
What was Hollywood thinking with this movie! I hated, hated, hated it. BORING! I went afterwards and demanded my money back. They refused.	*polarity of -1 with magnitude of 3.3*

VADER:

Phrase	***Result***
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006.	*{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}*
July wheat is expected to fall by nearly 14 percent to $4.00 per bushel, the lowest since September 2006.	*{'neg': 0.126, 'neu': 0.874, 'pos': 0.0, 'compound': -0.3818}*
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006, while soybean November is seen tumbling 17 percent to $7.60 per bushel, the lowest since May 2007.	*{'neg': 0.069, 'neu': 0.931, 'pos': 0.0, 'compound': -0.3818}*
`Bladerunner` is often touted as one of the best science fiction films ever made. Indeed, it satisfies many of the requisites for good sci-fi: a future world with flying cars and humanoid robots attempting to rebel against their creators. But more than anything, `Bladerunner` is a fantastic exploration of the nature of what it means to be human. If we create robots which can think, will they become human? And if they do, what makes us unique? Indeed, how can we be sure we're not human in any case? `Bladerunner` explored these issues before such movies as `The Matrix,' and did so intelligently. The visual effects and score by Vangelis set the mood. See this movie in a dark theatre to appreciate it fully. Highly recommended!	*{'neg': 0.02, 'neu': 0.774, 'pos': 0.206, 'compound': 0.9794}*
What was Hollywood thinking with this movie! I hated, hated, hated it. BORING! I went afterwards and demanded my money back. They refused.	*{'neg': 0.575, 'neu': 0.425, 'pos': 0.0, 'compound': -0.9653}*

Linguistic Inquiry and Word Counts:

Phrase	LIWC Positive Score	LIWC Negative Score
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006.	0	0
July wheat is expected to fall by nearly 14 percent to $4.00 per bushel, the lowest since September 2006.	0	5
July wheat is expected to rise by nearly 14 percent to $4.00 per bushel, the highest since September 2006, while soybean November is seen tumbling 17 percent to $7.60 per bushel, the lowest since May 2007.	0	2.7
`Bladerunner` is often touted as one of the best science fiction films ever made. Indeed, it satisfies many of the requisites for good sci-fi: a future world with flying cars and humanoid robots attempting to rebel against their creators. But more than anything, `Bladerunner` is a fantastic exploration of the nature of what it means to be human. If we create robots which can think, will they become human? And if they do, what makes us unique? Indeed, how can we be sure we're not human in any case? `Bladerunner` explored these issues before such movies as `The Matrix,' and did so intelligently. The visual effects and score by Vangelis set the mood. See this movie in a dark theatre to appreciate it fully. Highly recommended!	5.5	0.8
What was Hollywood thinking with this movie! I hated, hated, hated it. BORING! I went afterwards and demanded my money back. They refused.	0	16.7

Results

As shown in the results, all sentiment analysis tools fail to capture economic sentiment in a consistent way.

Google NLP gives the same ratings to all economic phrases, without any distinction between phrases 1, 2 and 3.
On the other hand VADER captures some difference between phrase 1, phrases 2 and 3, but it doesn't give any weight in terms of positivity/negativity to phrase 1. Moreover phrase 3 is less negative than phrase 2 because it gets the smaller than phrase 2 negative soybeans sentiment, without giving any weight to the positive wheat-related sentiment. If wheat-related sentiment would have been evaluated as positive, there would have been some pollution.
Also LIWC fails. Even if for phrases 1, 2 and 3 the dictionary was set to 'Scientific', the instrument fails to capture the positive economic sentiment in phrase 1 while captures negativity in phrase 2. In phrase 3, soybean negative sentiment is added to positive wheat-related sentiment (that is 0, because it's phrase 1). In a certain way it behaved as VADER.

Conversely all the three instruments are quite sharp in measuring movie reviews sentiment.
This is given by the fact that almost all previous research involved in the lexicons/dictionaries for sentiment analysis was developed using reviews taken from the web. For this reason many sentiment analyzers are usually more focused on measuring sentiment through the use of adjectives, emoticons and also slang sentences, while they are insensitive to signs as "+" or "–" and they don't fit well if used on neutral statements.
Since they are focused in measuring reviews sentiment, many researchers gave to objectivity the meaning of lack of sentiment: SentiWordNet for example sums to 1 all positivity, negativity and objectivity scores, with the result that a 100% objective text cannot be neither positive nor negative.
This is also confirmed by the fact that VADER lexicon doesn't contain any symbol as "+", "-" or words as "rise", "fall" and "decrease" (while "increase" is present and positive), but it contains almost all the emoticons. Moreover VADER at least doesn't consider objectivity as lack of sentiment, since the *compound score* is a function of *positivity*/*negativity* and *neutrality*.
Moreover sentiment context is important because it can change a word own sentiment polarity: for example the word 'unpredictable' is positive if referred to a movie plot, buti s negative in a price time series context. A lexicon built on movie reviews will give to "unpredictable" a positive polarity, while a lexicon based on economic statements will give to the same word a negative polarity.

Having said that, it's therefore strongly suggested to build a modified dictionary that mirrors our purposes, namely to measure economic sentiment in news.
In order to do so we should create a lexicon that contains words peculiar to agricultural market and economics and, for each word, sentiment positivity and negativity measurements have to be given. In doing so, we should also take into account the fact that newspaper articles are objective/neutral, therefore it's advisable to fill the dictionary with neutral words giving an "*economic sentiment*" measure to them.

Since building a lexicon can be a time-consuming activity, a work-around can be to extrapolate from each article only sentiment-bearing sentences and to build a lexicon only on that set.

For example phrase 3 contains sentiment related to two different objects: wheat and soybeans. In this context the changing of object can be detected by the use of the conjunction "while", that, when is not used in the sense of "in the meantime", means "although" and therefore signs a change of subject. Such words are usually part of a lexicon and they are used in the dictionary to calibrate sentiment measures, as in the case of negations. By using these words it's possible to deliver a synctatic analysis that permits to minimize the set of phrases that have to be analyzed during the sentiment analysis and, also considering the context of neutral newspaper articles, it permits also to minimize the lexicon needed.

Once the lexicon is done, an articles-oriented natural language model can be built in order to refine sentiment measurements by taking into account the peculiar structure of agriculture-related newspaper articles.

Lexicon Analysis

Existing lexicons are:

Analysis Tool	***Website***	***Lexicon***
Subjectivity lexicon MPQA	*http://mpqa.cs.pitt.edu*	*http://mpqa.cs.pitt.edu/lexicons/*
SentiWordNet	*http://sentiwordnet.isti.cnr.it/*	*http://www-3.unipv.it/wnop/* *http://www.unipv.it/wnop/micrownop.tgz*
Harvard General Inquirer	*http://www.wjh.harvard.edu/~inquirer/*	Roger Hurwitz, **rhhu@csail.mit.edu**
Bing Liu's Lexicon	*https://www.cs.uic.edu/~liub/*	*http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar*
Linguistic Inquiry and Word Counts (Proprietary)	*http://liwc.wpengine.com/*	*It can be tried at its own webpage.* *It can be tuned for personal writing, scientific writing et cetera.*
VADER Sentiment Analysis	*https://github.com/cjhutto/vaderSentiment*	*https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader_sentiment_lexicon.txt*
Google NLP	*https://cloud.google.com/natural-language/*	*Under Research*

Amongst all these lexicons it has to be noted that at the moment *Harvard General Inquirer* is not reachable and Roger Hurwitz has to be contacted with questions about access to the General Inquirer (**rhhu@csail.mit.edu**).

SentiWordNet considers positive/negative sentiment as something opposed to objectivity. In fact in its lexicon (based on *WordNet*) only positivity and negativity values are provided. However it's a good starting point for a lexicon because it has many numerical positivity/negativity values. SentiWordNet is included in *NLTK* python package, but probably it will deliver many objective scores.

Subjectivity lexicon MPQA is a lexicon for subjectivity measures, namely for each word it's accounted its subjectivity type (weakly subjective or strongly subjective), the Part Of Speech that represents (verb, adjective et cetera), if it's stemmed or not and a qualitative assessment about its polarity (positive or negative).
Unfortunately this lexicon hasn't quantitative sentiment polarity values.

Bing Liu's Lexicon is composed by just two lists: one list contains positive words while the other one contains negative words. Unfortunately his work lacks of any other word-related additional information, therefore just each word qualitative polarity is present. However on Bing Liu's website can be found a few research papers on the subject and they are certainly of our interest since they deal in a very detailed way with sentiment-extraction.

Linguistic Inquiry and Word Counts is a proprietary lexicon and it has been created following these steps:

- Words collection: words are collected from dictionaries, then classes are created by a few judges.
- Judge Rating Phase: each word is inserted in a class in order to build conceptually homogeneous word classes. The principle is "goodness of fit" and the process is performed by a group of 4-8 judges.
- Base Rate Analyses: using twitter tweets, facebook posts, blogs et cetera, words appearance in different context is checked, then the least representative classes are dropped off.
- Candidate Word List Generation: high frequency words not included in pre-existing classes are included in dictionary categories if they show some degree of correlation with the classes.
- For each class it's computed a consistency estimate for each word (using Cronbach Alpha). Words that are under a certain consistency threshold are dropped off from final dictionary.
- All the steps are repetead to check for errors.

At the end of the process the lexicon is done and sentiment values are given.

VADER Sentiment Analysis tool is tested above and, by looking at its lexicon, it appears that it's well suited for twitter tweets: it has a detailed list of emoticons, but for example it lacks of the verb "to fall", the word "price" and the adjective "decreasing". Probably all these words, being not accounted, are no-sentiment-bearing words, therefore VADER scores are not accurate in an agricultural newspaper context. To be clear: if we think to the Tri-Gram "Price is decreasing", VADER will give 0 positive as 0 negative sentiment and 1 as neutral sentiment score (tested), while in economic terms this is not true since the sentence "Price is decreasing" is full of *economic sentiment*.

At the moment ***Google NLP*** lexicon is still under research, but probably if used in our context it suffers of the same flaws of the other lexicons given the above results.

Conclusion

It has been shown that lexicons are *context-dependent*, therefore we need a lexicon suited to our context.
Unfortunately most of the open-source research points to different contexts, such as reviews and tweets, therefore a big part of the lexicons, expecially the open-source ones, are *human-sentiment oriented*. Being *human-sentiment oriented* means that they are quite insensitive to economic statements. This fact produces erroneous sentiment measurements that are not correlated to prices.

It can happen that, if used on real articles, these instruments will get more accurate sentiment estimates by the effect of the whole articles wording. However this fact may be countered by sentiment pollution caused by other objects mentioned in the articles as articles context that may distort each word sentiment polarity.

In order to solve this issue a sentiment-ratings-comprehensive lexicon has to be created for measuring *economic sentiment*. Once the lexicon is finished, synctatic structures have to be modeled to calibrate sentiment extraction by taking into account the many words interactions that may distort a sentence sentiment from just the sum of each sentence-contained word sentiment.
Since this can be a time-consuming task, it's proposed to extract from the articles just sentiment-bearing sentences regarding one commodity (wheat for example). By doing so sentiment noise caused by other article's objects is removed and the lexicon needed is minimized by minimizing the context that it will face.

Stressing this concept we can think of extracting not only sentences related to one commodity, but also related to only one feature. For example it can bring benefits to extract just articles sentences that deal with wheat price and then build a lexicon on them.
In this way the lexicon will be minimized to only the "wheat price" context, all other sentiment noise caused by other objects will be removed and probably the price correlation will be found.